IN THE UNITED STATES DISTRICT COURT FOR

THE DISTRICT OF MASSACHUSETTS

NUANCE COMMUNICATIONS, INC.,

Plaintiff and Counterclaim Defendant,

v.

OMILIA NATURAL LANGUAGE SOLUTIONS, LTD.,

Defendant and Counterclaim Plaintiff.

Case No. 1:19-cv-11438-PBS

NUANCE COMMUNICATIONS, INC.'S STATEMENT OF FACTS AND RESPONSE TO OMILIA NATURAL LANGUAGE SOLUTIONS, LTD.'S STATEMENT OF PURPORTED MATERIAL FACTS IN SUPPORT OF ITS MOTION FOR PARTIAL SUMMARY JUDGMENT FOR THE INVALIDITY OF U.S. PATENT 6,999,925

Pursuant to Local Rule 56.1 Plaintiff and Counterclaim Defendant Nuance Communications, Inc. ("Nuance") hereby submits this response to Defendant and Counterclaim Plaintiff Omilia Natural Language Solutions, Ltd.'s ("Omilia") Statement of Purported Material Facts in Support of its Motion for Partial Summary Judgment for the Invalidity of U.S. Patent 6,999,925.

I. RESPONSE TO OMILIA'S STATEMENT OF FACTS

1. The '061 patent, titled "Method and System for Generating Squeezed Acoustic Models For Specialized Speech Recognizer," issued from an application filed on August 14, 2000, that claims priority to a Foreign European Patent Application No. 99116684, filed on August 25, 1999. *See* Ex. 1 to the Declaration of Hallie Kiernan ("Kiernan Decl.," DN 211), ("'061 patent"), at [22], [30], [54].

Response: Undisputed.

2. The '061 patent names as inventors Volker Fischer, Siegfried Kunzmann, and Claire Waast-Ricard. '061 patent at [75]; *see also* Kiernan Decl. Ex. 2.

Response: Undisputed.

3. The inventors assigned the '061 patent to International Business Machines Corporation ("IBM") in September 2000. '061 patent at 1, [73].

Response: Undisputed.

4. The '061 patent issued to IBM on September 7, 2004. '061 patent at [45].

Response: Undisputed.

5. The '061 patent claims a method and apparatus for automatically generating or adapting a second speech recognizer from a first speech recognizer. *See, e.g.,* '061 patent, Abstract, cl. 6 and 15.

Response: Disputed. Omilia does not properly quote any claim language from the '061 patent and mischaracterizes claim terms; Nuance objects to Omilia's use of claim terms out of context as misleading; the term "adapting" does not even appear in claims 6 and 15

of the ’061 patent; Nuance objects to Omilia’s statement as immaterial and irrelevant.

6. The ’061 patent would have expired on or around January 31, 2021, but it instead expired on September 7, 2008 due to IBM’s failure to pay maintenance fees. *See* ’061 patent at [22], [*]; 35 U.S.C. § 154; Kiernan Decl. Ex. 3.

Response: Undisputed.

7. The ’925 patent, titled “Method and Apparatus for Phonetic Context Adaptation For Improved Speech Recognition,” issued from an application filed on November 13, 2001, that claims priority to Foreign European Patent Application No. 00124795, filed on November 14, 2000. Kiernan Decl. Ex. 4 (“’925 patent”) at [22], [30], [54].

Response: Undisputed.

8. The ’925 patent names as inventors Volker Fischer, Siegfried Kunzmann, Eric-W. Janke, and Jon Tyrrell. ’925 patent at [75]; *see also* Kiernan Decl. Ex. 5.

Response: Undisputed.

9. Volker Fischer and Siegfried Kunzmann are also inventors of the ’061 patent. *Compare* ’925 patent at [75] with ’061 patent at [75]; *see also* Kiernan Decl. Exs. 2 & 5.

Response: Undisputed that Volker Fischer and Siegfried Kunzmann are named inventors of the ’061 patent; however, Nuance objects to this statement as misleading. The ’925 patent has two other inventors not listed as inventors of the ’061 patent, who made patentable contributions to the ’925 patent.

10. The inventors assigned the ’925 patent to IBM in October 2001. *See* Kiernan Decl. Ex. 6 (NUANCE0000000412).

Response: Undisputed.

11. The ’925 patent issued to IBM on February 14, 2006. ’925 patent at [45].

Response: Undisputed.

12. The ’925 patent is not subject to a terminal disclaimer. *See* ’925 patent at 1, [*].

Response: Undisputed that the ’925 patent is not currently subject to a terminal disclaimer; Nuance objects to Omilia’s statement as misleading and irrelevant; the ’925

patent should not be subject to any terminal disclaimer as it is a patentably distinct invention.

13. The ’925 patent also claims a method and system for generating or adapting a second speech recognizer from a first speech recognizer. *See, e.g.,* ’925 patent, Abstract, cl. 1, 14, and 27.

Response: Disputed. The ’925 patent does not “also” claim a method and system for generating or adapting a second speech recognizer from a first speech recognizer because the ’061 patent does not claim a method and system for generating or adapting a second speech recognizer from a first speech recognizer—Omilia does not properly quote any claim language from the ’061 patent and mischaracterizes claim terms; Nuance objects to Omilia’s use of claim terms out of context as misleading; the term “adapting” does not even appear in claims 6 and 15 of the ’061 patent; Omilia does not properly quote any claim language from the ’925 patent and mischaracterizes claim terms; Nuance objects to Omilia’s use of claim terms out of context as misleading; Nuance objects to Omilia’s statement as immaterial and irrelevant.

14. The ’061 patent claims were not evaluated as an obviousness type double patenting reference during the prosecution of the ’925 patent. ’925 patent at 1, [56].

Response: Disputed. The Examiner expressly considered the ’061 Patent; in an October 19, 2020 filing, the Examiner’s search strategy and results show that the Examiner specifically searched for “6789061” (the ’061 Patent’s full number) in the U.S. Patents database. Richey Decl. Ex. 1 (’925 Patent File History); Thus, the examiner considered the ’061 Patent, and did not enter an ODP rejection or a §§ 102/103 rejection; he is presumed to have done his job in declining to take such action, and any consideration of the validity of the ’925 Patent in view of the ’061 Patent would have implicated an ODP analysis. Moreover, the European patent application (EP99116684.4) that led to the ’061 Patent, was cited, incorporated by reference, and distinguished in the specification of the ’925 Patent, identifying key distinctions between ’061 Patent’s disclosures and the

invention claimed in the ’925 Patent. Dkt. 211-4 (’925 Patent) at 6:57-7:17, 8:61-9:3.

15. The ’925 patent was assigned to Nuance Communications, Inc. when it purchased the ’925 patent from IBM, along with other patents on December 31, 2008. *See* ECF No. 84-2 at 10.

Response: Undisputed.

16. The ’925 patent will expire on or around October 12, 2023. See ’925 patent at 1, [22], [*]; 35 U.S.C. § 154.

Response: Undisputed.

17. United States Patent No. 6,912,499, titled “Method And Apparatus For Training A Multilingual Speech Model Set,” (“Sabourin”), issued from an application filed on August 31, 1999, and is prior art under 35 U.S.C. § 102(e). *See* Kiernan Decl. Ex. 7, at [22], [54].

> **Response:** Undisputed that United States Patent No. 6,912,499 is titled “Method And Apparatus For Training A Multilingual Speech Model Set,” and issued from an application filed on August 31, 1999. Disputed as to Sabourin being prior art; Sabourin is not prior art to the ’061 patent, which has a priority date of August 25, 1999. *See* Omilia’s Statement of Facts #1; Ex. 1 to Kiernan Decl. at [22], [30], [54].

18. Nuance’s expert, Karen Livescu, in a declaration submitted in support of Nuance’s claim construction on the issue of the meaning of a “second language” in the claims of the ’925 patent stated that “[a] POSITA would understand that based on this language, both the first domain and the second domain can include multiple languages, but need not do so.” Dkt. No. 86 at ¶ 36.

> **Response:** Undisputed that Omilia has correctly quoted a passage from Karen Livescu’s declaration; Nuance objects to Omilia’s statement as an immaterial fact and irrelevant. The Court construed “second language” in its claim construction order; thus, any underlying facts as to the meaning of this term have been considered by the Court. Nuance further objects to Omilia’s use of this excerpt out of context as misleading.

19. Multilingual recognizers that recognized two or more languages were known by November 14, 2000, the earliest priority date on the face of the ’925 patent. Declaration of Jordan Cohen at ¶ 85 (and the materials cited therein).

Response: Nuance is unable to discern what Omilia means by its vague statement that multilingual recognizers that “recognize two or more languages” were known by November 14, 2000, and therefore disputes it, as presented here, as a standalone assertion; Nuance objects to Omilia’s statement as immaterial, irrelevant, and out of context. Declaration of Chris Schmandt in Opposition to Defendant’s Motion for Partial Summary Judgment for the Invalidity of the ’925 Patent (“Schmandt Decl.”) ¶¶ 146-153.

II. ADDITIONAL DISPUTED MATERIAL FACTS REQUIRING DENIAL OF PARTIAL SUMMARY JUDGMENT

20. While the construction of patent claims is a question of law for a judge to determine, claim construction may involve underlying factual disputes. Claim terms are generally given their ordinary and customary meaning, which is the meaning that the term would have to a person of ordinary skill in the art in question at the time of the invention. According to the declaration of Omilia’s expert, Dr. Cohen, the following terms in claim 1 of the ’925 Patent somehow correspond to different language in claim 6 of the ’061 Patent, thus requiring construction before differences or equivalences can be identified: “acoustic model” (DN 210 ¶¶ 47-50, 59); “decision network” (*id.*); “phonetic contexts” (*id.*); “adapted to” (*id.* ¶ 57); “domain” (*id.* ¶¶ 55-56); “re-estimating the first decision network and said corresponding phonetic contexts” (*id.* ¶¶ 58, 60, 62); “second acoustic model” and “second decision network” (*id.* ¶¶ 59, 60); “based on domain-specific training data” (*id.* ¶ 63); “utilize a decision tree” (*id.* ¶ 67, 69); “number of nodes in the second decision network is not fixed by the number of nodes in the first decision network” (*id.* ¶ 71); and “partitioning said training data using said first decision network of said first speech recognizer” (*id.* ¶¶ 72-74). To the extent these claim terms require construction, any underlying factual issues are disputed. *See* Schmandt Decl. ¶¶ 69-79.

21. Further, Dr. Cohen identifies several additional terms in claim 27 of the '925 Patent as corresponding to language in the '061 patent, thereby also requiring construction: "receiving domain specific training data of a second domain" (DN 210 ¶ 81; *see also id*. ¶¶ 55-56); and "wherein the first domain comprises at least a first language, wherein the second domain comprises at least a second language, and wherein the second speech recognizer is a multi-lingual speech recognizer." (*Id*. ¶¶ 82-94) To the extent these claim terms require construction, any underlying factual issues are disputed. *See* Schmandt Decl. ¶¶ 143-152.

22. Furthermore, according to Dr. Cohen, the following additional terms in claim 6 of the '061 Patent are the terms to which language in claim 1 of the '925 Patent allegedly corresponds, meaning that these terms must also be construed: "set of states" (DN 210 ¶¶ 51-53, 61, 64-66, 68, 69); "set of probability density functions" (*id*. ¶ 51); "distinctive of" (*id*. ¶ 57); "application" (*id*. ¶¶ 55-56); "selecting a subset of states" (*id*. ¶¶ 59, 61, 62, 71); "exploits application specific training data" (*id*. ¶ 63); and "associating a multitude of speech frames of the training data with the correct states of the first speech recognizer" (*id*. ¶¶ 72-74). To the extent these claim terms require construction, any underlying factual issues are disputed. *See* Schmandt Decl. ¶¶ 69-79.

23. The primary purpose of the '061 Patent's invention is to minimize the resources required of a speech recognizer, so it can run on a less powerful computational platform and/or with a smaller memory footprint. *See* in general col. 1:66 – 2:35. The patent describes a way to determine which parts of the acoustic model of an initial, general purpose, recognizer are required for adequate performance on a specific task or domain, and then remove the parts which do not contribute much to recognition performance, to obtain a smaller recognizer, which may be important in circumstances where the general purpose recognizer would require too many resources (e.g., memory and/or computing power). Schmandt Decl. ¶ 54.

24. This is referred to as "squeezing" and "pruning" the recognizer, i.e. removing, or subsetting, some portions of the acoustic model which are less appropriate for a specific and limited application, as opposed to the more general purpose speech recognizer. Specifically, this

is done by starting with the set of states of the general recognizer, then choosing a subset of Hidden Markov Model ("HMM") states – squeezing – and pronunciation probability density functions ("PDFs") – pruning – associated with some of those states (col. 2:15-25), based on which of those HMM states and PDFs are most likely to have produced each frame of the application-specific speech training data (col. 7:3-8). This process does not have the ability to change the phonetic contexts of the general purpose speech recognizer (and in fact it has no access to the phonetic context decision tree). It can only delete some of the states, which had been previously derived by sorting acoustic training data using the decision tree based on their phonetic contexts. Schmandt Decl. ¶ 55.

25. To squeeze HMM states, the training data is "Viterbi aligned" with the set of HMM states. Viterbi alignment is a process to determine which of the states would most likely produce a correct transcription, on a frame-by-fame basis. Those states which appear frequently after alignment with the training data are more important to correct recognition and must be retained. Those states which are seen infrequently may be subject to squeezing (col. 6:8-33). Schmandt Decl. ¶ 56.

26. Similarly, the sets of PDFs can be pruned by comparing the acoustic outputs that each surviving state produces versus each matched acoustic frame. Since the GMMs (Gaussian Mixture Models) describe a mixture of base output probability PDFs, it is possible to determine how similar the various PDFs in the mixture are to the actually recorded acoustic features, on a frame by frame basis. If the training data contains speech which maps well to only one of the acoustic outputs, the other PDFs may be pruned (col. 6:34-51). Put another way, if an HMM state can, via its acoustic realization PDFs, cover several ways that particular phonemic context may be spoken, but only one of those ways of speaking is encountered in the training data, then the remaining (non-encountered) ways of speaking can be removed. Schmandt Decl. ¶ 57.

27. The '925 Patent teaches an improved method for efficiently refining a speech recognizer to better handle particular domains (such as languages, dialects, task areas, or user-specific information, *see* col. 6:3-8, 9:6-23). At the time of the claimed invention of the '925

Patent, general purpose, large vocabulary, continuous speech (the way humans speak when talking normally, and a more difficult type of speech for recognizers to process as compared to words spoken with pauses between each word) recognizers were known in the art and were in general usage. However, while such speech recognizers provide good overall speech recognition quality, when a use-case centers on a specific domain – which may differ from the domain of the general-purpose speech recognizer in terms of its vocabulary, dialect, acoustic conditions, and so on – the speech recognition accuracy may not be as high as would be possible with a more specialized speech recognizer. Schmandt Decl. ¶ 59.

28. While additional domains may be similar or related to the general-purpose domain, there may be particular phonetic contexts—particular ways that words are pronounced, or that words or sub-word units are used in sequence together—that are characteristic of the domain. In such cases, taking domain-specific data into consideration can improve the accuracy of a speech recognizer for a particular context of use. But using the invention of the '925 Patent, there is not a need to completely train the new speech recognizer from scratch, particularly if there are parts (even, perhaps, substantial parts) of the original speech recognizer that can be used in combination with a (smaller) set of specialized data to adapt the recognizer to the new domain. Schmandt Decl. ¶ 60.

29. Prior art methods of providing domain-specific speech recognizers tended to entail either merely "select[ing] a domain specific subset from the phonetic context inventory" of the general-purpose recognizer, or extensively-retraining a speech recognizer, based on a large, specialized dataset. *See generally* col. 1:15 - 2:17, quotation from col. 2:1-3. By contrast, the inventions claimed in the '925 Patent start with a general purpose, large vocabulary speech recognizer, and by adding only a modest amount of additional specialized training data ("adaptation data," *see* col. 2:56) to the initial dataset, are able to efficiently provide a new, specialized speech recognizer that includes the new training data and new phonetic contexts. Because the collection of training data and subsequent training of a speech recognizer is

expensive and time consuming, adapting a general purpose speech recognizer to a specific domain greatly reduces development costs and time to market. Schmandt Decl. ¶ 61.

30. The technique disclosed in the ’925 Patent is to start with the original decision network and efficiently recalculate it via the use of a smaller amount of domain-specific training data. The data is processed through the decision network in the same way as the original training data was processed to produce the original recognizer, and the new data is allowed to modify the decision tree in the process. The split-and-merge methods described for both the original and new adaptation data control the decision tree’s topology. In the process of keeping a balance while partitioning the training data – separating it into the various leaf nodes of the tree to provide audio frame material to create the acoustic model – nodes may be added or removed at any level in the decision tree. Once all the adaptation data has been processed and a new phonemic decision tree thereby created, the acoustic model can be generated. Schmandt Decl. ¶ 62.

31. The inventors of the ’925 Patent acknowledge and distinguish that patent from the techniques of the ’061 Patent. Repeatedly throughout the ’925 Patent, the specification discusses and distinguishes European patent application EP 99116684.4, which the ’061 Patent indicates on its face is the application from which the ’061 Patent claims priority. Schmandt Decl. ¶ 63.

32. Starting at col. 6:32, the inventors describe the training procedure of a speech recognizer as a two stage process, “1.) the determination of *relevant* acoustic contexts and 2.) the estimation of acoustic model parameters.” (Col. 6:34-36 (emphasis added). They then describe how some existing approaches (MAP, MLLR) focus exclusively on re-estimation of the acoustic model parameters, “Importantly, these approaches leave the phonetic contexts unchanged.” (Col. 6:44). At this point, the inventors have articulated two distinct approaches, and the ’925 Patent’s invention will be seen (in the rest of the specification) to correspond to the first. They point out that the second set of techniques have heretofore been mostly used to adapt speaker independent recognizers to a particular user (col. 6:47-49). Schmandt Decl. ¶ 64.

33. The inventors go on (col. 6:50) to discuss additional techniques of modifying acoustic models to achieve better domain specific speaker independent recognition. Here they cite to EP 99116684.4 (from which the '061 Patent claims priority) describing it as "selecting a subset of probability density functions (PDFs) being distinctive for the domain." (Col. 6:63-64). Schmandt Decl. ¶ 65.

34. They specifically distinguish the '925 Patent invention from the EP invention, characterizing the two approaches as "orthogonal[] to" each other, because the '925 Patent invention "focuses on re-estimation of phonetic contexts or – in other words – the adaptation of the recognizer's sub-word inventory to a special domain." Col. 6:66 - 7:2. This phonetic context is what was called out in part 1) (above) of the description of the bifurcation of two major approaches at col. 6:32. As the inventors state here, the '925 Patent recalculates the phonetic context, while the '061 Patent has nothing to do with it. Schmandt Decl. ¶ 66.

35. Column 7 goes on to explain clearly how the invention of the '925 Patent is different: "Whereas in any speaker adaptation algorithm, as well as in the above mentioned documents of *V. Fischer et al.,* the phonetic contexts once estimated by the training procedure are fixed, the present invention utilizes a small amount of upfront training data for the domain specific insertion, deletion, or adaptation of phones in their respective context. Thus re-estimation of the phonetic contexts refers to a (complete) recalculation of the decision network and its corresponding phonetic contexts based on the general speech recognizer decision network. This is considerably different from just 'selecting' a subset of the general speech recognizer decision network and phonetic contexts or simply 'enhancing' the decision network by making a leaf node an interior node by attaching a new sub-tree with new leaf nodes and further phonetic contexts." (Col. 7:2-17). Schmandt Decl. ¶ 67.

36. To be clear, the inventors themselves describe above how the '925 and '061 Patents differ. In no uncertain terms, they state that the '925 Patent's invention "is considerably different" from the '061 Patent, and that the two patents describe two distinct approaches to

solving two different problems in computerized speech recognition. (Col. 7:12). Schmandt Decl. ¶ 68.

37. A POSITA would understand that the claim language of claim 6 of the ’061 Patent identified in the table below is different from the claim language of claim 1 of the ’925 patent identified in the table below, for the reasons set forth in Schmandt Decl. ¶¶ 80-98.

Claim 6 of ’061 Patent	Claim 1 of ’925 Patent
wherein the first speech recognizer includes a set of states and a set of probability density functions assembling output probabilities for an observation of a speech frame in the states,	said first speech recognizer comprising a first acoustic model with a first decision network and corresponding first phonetic contexts

38. A POSITA would understand that the claim language of claim 6 of the ’061 Patent identified in the table below is different from the claim language of claim 1 of the ’925 patent identified in the table below, for the reasons set forth in Schmandt Decl. ¶¶ 99-108.

Claim 6 of ’061 Patent	Claim 1 of ’925 Patent
generating, from the set of states of the first speech recognizer, a set of states of the second speech recognizer by selecting a subset of states of the first speech recognizer . . . and generating, from the set of probability density functions of the first speech recognizer, a set of probability density functions of the second speech recognizer by selecting a subset of probability density functions of the first speech recognizer being distinctive of the particular application, such that the second speech recognizer is at least one tailored to the particular application and requires reduced resources compared to the first speech recognizer.	based on said first acoustic model, generating a second acoustic model with a second decision network and corresponding second phonetic contexts for said second speech recognizer by re-estimating said first decision network and said corresponding first phonetic contexts

39. A POSITA would understand that the claim language of claim 6 of the ’061 Patent identified in the table below is different from the claim language of claim 1 of the ’925 patent identified in the table below, for the reasons set forth in Schmandt Decl. ¶¶ 109-112.

Claim 6 of ’061 Patent	Claim 1 of ’925 Patent
wherein selecting at least one of the subset of states and the subset of probability density functions of the first speech recognizer exploits application specific training data.	based on domain-specific training data,

40. A POSITA would understand that the claim language of claim 6 of the ’061 Patent identified in the table below is different from the claim language of claim 1 of the ’925 patent identified in the table below, for the reasons set forth in Schmandt Decl. ¶¶ 113-122.

Claim 6 of ’061 Patent	Claim 1 of ’925 Patent
wherein the first speech recognizer includes a set of states and a set of probability density functions assembling output probabilities for an observation of a speech frame in the states,	wherein said first decision network and said second decision network utilize a phonetic decision [t]ree to perform speech recognition operations

41. A POSITA would understand that the claim language of claim 6 of the ’061 Patent identified in the table below is different from the claim language of claim 1 of the ’925 patent identified in the table below, for the reasons set forth in Schmandt Decl. ¶¶ 123-126.

Claim 6 of the ’061 Patent	Claim 1 of the ’925 Patent
generating, from the set of states of the first speech recognizer, a set of states of the second speech recognizer by selecting a subset of states of the first speech recognizer . . . generating, from the set of probability density functions of the first speech recognizer, a set of probability density functions of the second speech recognizer by selecting a subset of probability density functions of the first speech recognizer	wherein the number of nodes in the second decision network is not fixed by the number of nodes in the first decision network

42. A POSITA would understand that the claim language of claim 6 of the ’061 Patent identified in the table below is different from the claim language of claim 1 of the ’925 patent identified in the table below, for the reasons set forth in Schmandt Decl. ¶¶ 127-130.

Claim 6 of the '061 Patent	Claim 1 of the '925 Patent
wherein selecting the subset of states comprises associating a multitude of speech frames of the training data with the correct states of the first speech recognizer	and wherein said re-estimating comprises partitioning said training data using said first decision network of said first speech recognizer

43. The differences between claim 6 of the '061 Patent and claim 1 of the '925 Patent render claim 1 of the '925 Patent patentably distinct. Schmandt Decl. ¶¶ 131-139.

44. When comparing claim 6 of the '061 Patent and claim 1 of the '925 Patent as a whole, claim 1 of the '925 Patent is patentably distinct. Schmandt Decl. ¶ 140.

45. The substantive terms of '925 Patent claim 14 (a dependent claim that includes the limitations of claims 10, 11, 13, 14 and 15) are patentably distinct from '061 Patent claim 15 for the same reasons that claim 1 of the '925 Patent is patentably distinct from claim 6 of the '061 Patent. Schmandt Decl. ¶ 141; *id.* ¶¶ 80-140.

46. '925 Patent claim 27 is patentably distinct from '061 Patent claim 6 for the same reasons as claim 1. Schmandt Decl. ¶¶ 145-177; *id.* ¶¶ 80-140.

47. The final limitation of claim 27, "wherein the first domain comprises at least a first language, wherein the second domain comprises at least a second language, and wherein the second speech recognizer is a multi-lingual speech recognizer," is simply absent from claim 6 of the '061 Patent; therefore, this limitation, and claim 27 as a whole, are not an obvious variant of '061 Patent claim 6. Schmandt Decl. ¶ 145; *compare* '061 Patent claim 6 *with* '925 Patent claim 27.

48. The Schultz paper attached to Dr. Cohen's declaration as Exhibit D, "Multilingual and Crosslingual Speech Recognition," apparently to demonstrate that multilingual recognizers were known in the art. While perhaps true, Dr. Cohen has failed to demonstrate that any concept in this publication mapped to the invention disclosed in the '925 Patent. Further, in conducting the ODP analysis, it is not permissible to combine the reference patent with other prior art. In fact, the paper discusses multilingual recognizers which are designed up front for multiple languages, based on creating an entirely new recognizer by initially combining training data from

multiple languages which have been pre-selected to give broad phoneme coverage. This does not correspond to the invention claimed in the ’925 Patent, and is different from the ’925 Patent process of re-estimation to augment a first recognizer (that includes a first language) with a second domain that ultimately results in a multi-lingual recognizer. The paper also discusses “crosslingual” recognition in which “the developed multilingual systems are applied to recognize new unseen languages without any additional training” (Schultz Exhibit D at p. 5, emphasis original). The ’925 Patent is clearly a different technique which relies on additional training data. *See* Schmandt Decl. ¶¶ 153.

49. The differences of claim 27 of the ’925 Patent and claim 6 of the ’061 patent are patentably distinct over the Schultz paper attached to Cohen’s declaration as Exhibit D, the Schultz paper attached to Cohen’s declaration as Exhibit E, the Schultz paper attached to Cohen’s declaration as Exhibit G, and Sabourin reference attached to Cohen’s declaration as Exhibit F. Schmandt Decl. ¶¶ 153-171.

50. The Sabourin patent, attached to Cohen’s declaration as Exhibit F, is completely distinct from claim 27 of the ’925 Patent. Sabourin does not contribute in any way to the claims of the ’061 Patent in comparing them to the ’925 Patent. Sabourin describes in detail a process for creation of a multi-lingual recognizer. It begins with letter to sound rules (found in neither the ’061 Patent nor the ’925 Patent, both of which use phonetically labelled or HMM aligned training data) which allow training data to be assigned to an apparently hand selected set of pre-determined phonemes for the joint languages. This rule set is not built computationally but rather by a trained linguist (col. 5:21-22). Phonemes in the second language are initially mapped to the nearest phonemes in the first language, not by a decision tree, but rather by evaluating similarity, based on a phonemic feature description of the phonemes (again, determined by a linguist or phonetician), finding which old phoneme is minimally different from the new one by minimizing a “transformation” rule, which basically compares which articulatory features are similar or different across the set of old phonemes. After this is done, the recognizer for the new language can be further adapted to the new pronunciations using maximum a posterior

adaptation (MAP), which is a process of updating HMM models with new data after they have been initially modeled by training data. But this technique is specifically taught away from in the ’925 Patent and the ’061 Patent. “Adaptation techniques known the (sic) within the state of the art, for example maximum a posteriori adaptation (MAP) or maximum likelihood linear regression (MLLR)… exclusively target the adaptation of the HMM parameters based on training data. Importantly, these approaches leave the phonetic contexts unchanged; that is, the decision network and the corresponding phonetic contexts are not modified by these technologies.” (’925 Patent at col. 6:36-46); (*see also* ’061 Patent at col. 4:29-34 (“The proposed solution according to the present invention to the above-mentioned problems is orthogonal to approaches which exploit speaker adaptation techniques, like, e.g., maximum a posteriori adaptation (MAP)…”). Eventually Sabourin does mention decision trees, but only in the context of splitting phonemes into allophones – *i.e.* variants of the phoneme with different acoustic realizations. But a separate HMM is not trained for each, but rather the allophone is adapted from the original HMM using, once again, MAP adaptation (col. 11:34). Schmandt Decl. ¶¶ 160, 162, 164-166, 171.

51. The Schultz paper submitted as Exhibit G to the Cohen declaration is different from both the ’061 Patent and the ’925 Patent. It is not evident how the doubling of some phonemes in this Schultz reference is in any way related to the deletion of phonemes (or states) in the ’061 Patent. Further, the doubling in Schultz is contradictory to the methods of the ’925 Patent, which maintain the richness of the original decision network while enhancing, or adding to it, with the new adaptive training data. There is no need to “make do” with existing contexts in the ’925 Patent, because its very goal is the ability to incorporate new contexts into the decision tree generation of the multilingual recognizer. Schmandt Decl. ¶ 170.

Date: January 7, 2021

Respectfully submitted,

/s/ Christian E. Mammen
WOMBLE BOND DICKINSON (US) LLP

David Greenbaum (*admitted pro hac vice*)
Nuance Communications, Inc.
1111 Macarthur Boulevard
Mahwah, NJ 07430
(857) 214-5524
david.greenbaum@nuance.com

Deborah M. Vernon (MA BBO No. 663937)
WOMBLE BOND DICKINSON (US) LLP
Independence Wharf
470 Atlantic Avenue, Suite 600
Boston, MA 02110
Telephone: (857) 287-3130
Deborah.Vernon@wbd-us.com

Christian E. Mammen (*admitted pro hac vice*)
Carrie Richey (*admitted pro hac vice*)
WOMBLE BOND DICKINSON (US) LLP
1841 Page Mill Road, Suite 200
Palo Alto, CA 94304
Telephone: (408) 341-3067
Chris.Mammen@wbd-us.com
Telephone: (408) 341-3060
Carrie.Richey@wbd-us.com

Christine H. Dupriest (*admitted pro hac vice*)
WOMBLE BOND DICKINSON (US) LLP
271 17th Street, Suite 2400
Atlanta, Georgia 30363
Telephone: (404) 962-7538
Christine.Dupriest@wbd-us.com

Attorneys for Plaintiff Nuance Communications, Inc.

CERTIFICATE OF SERVICE

I hereby certify that this document will be filed through the ECF system and will be sent electronically to the registered participants as identified on the Notice of Electronic Filing (NEF) and paper copies will be sent to those indicated as non-registered participants on January 7, 2021.

/s/ *Christian E. Mammen*
Christian E. Mammen